

Ім'я користувача:
Visnyk of the National Bank of Ukraine

ID перевірки:
1013254187

Дата перевірки:
09.12.2022 10:55:56 EET

Тип перевірки:
Doc vs Internet

Дата звіту:
09.12.2022 12:14:09 EET

ID користувача:
85646

Назва документа: 20210920_text

Кількість сторінок: 43 Кількість слів: 7322 Кількість символів: 52673 Розмір файлу: 1.38 MB ID файлу: 1012997815

3.18% Схожість

Найбільша схожість: 0.66% з Інтернет-джерелом (https://link.springer.com/chapter/10.1007/978-3-540-79005-1_7)

3.18% Джерела з Інтернету

143

Сторінка 45

Пошук збігів з Бібліотекою не проводився

0% Цитат

Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

0% Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи

10

ВСТУП

Починаючи з першого липня 2020 року НБУ почав упровадження реформи фінансового сектору, поширюючи вимоги прозорості, надійності та ефективності на небанківський фінансовий сектор. Першочерговою метою реформ є покращення якості послуг зі страхування та захист інтересів страхувальників.

Для ефективного нагляду, контролю та запровадження реформ небанківського фінансового ринку необхідне розуміння структури ринку та як ведуть свій бізнес його учасники. Наприклад, для різних бізнес-моделей притаманним є різний профіль ризику. Виділення однорідних груп зі схожими ризиками може сприяти більш детальному аналізу стійкості та платоспроможності страхових компаній, ефективному передбаченню кризових подій. Це дослідження має на меті зробити внесок саме у цьому напрямі. Матеріал присвячений виокремленню на основі кількісних показників бізнес-моделей українських страховиків та їх ключових характеристик. За нашими даними, це перша праця, що зачіпає таку тематику в Україні.

У ході роботи було проведено кластерний аналіз українського страхового ринку з метою визначення бізнес-моделей, що використовуються страховиками. Ми використали низку методів кластерного аналізу, до них увійшли ієрархічні, неієрархічні та метод машинного навчання. На основі методу K-середніх нам вдалось виділити 5 кластерів, що відповідають чотирьом бізнес-моделям – Корпоративна, Роздрібна, Універсальна (ділиться на два кластери) та Перестрахова. Перед застосуванням алгоритмів кластеризації було виділено штучний кластер “Неактивна”, (до нього увійшли компанії, що були малоактивними або взагалі не займалися страховою діяльністю, проте мали ліцензію і увійшли до вибірки). Оцінити якість проведеного нами дослідження допомагає ряд обрхованих коефіцієнтів, зокрема: коефіцієнт міграції та коефіцієнт силуету.

Бізнес-моделі було описано як за допомогою характеристик, на основі яких відбувалось розділення, так і використовуючи додаткові описові характеристики. Так, компанії бізнес-моделі Корпоративна страхують здебільшого юридичних осіб, моделі Роздрібна – навпаки, фізичних. Компанії з Універсальною бізнес-моделлю широко використовують офіси як канал продажу, тоді як з Перестраховою – не використовують їх взагалі. Детальніший опис кластерів наведено у тексті.

Далі ми показали як відбувалась міграція страховиків між кластерами у період з 2019 по 2020 рр. Значна міграція відбувалась між бізнес-моделями Перестрахова та Корпоративна та всередині бізнес-моделі Універсальна. Також ми спостерігали значні міграції до штучного кластеру Неактивні, тобто зупинку компаніями страхової діяльності. На основі цих міграцій можна емпірично зробити висновки про рівень ризику різних бізнес-моделей. Так, найбільша частка компаній, що вийшла з ринку за досліджуваний період спостерігалась у бізнес-моделі Перестрахова, більше половини компаній цього кластеру вийшли з ринку у 2020 р. Суттєва міграція до групи Неактивні також спостерігалась для бізнес-моделей Корпоративна та Роздрібна.

Робота має наступну структуру. У першому розділі подано огляд тематичної літератури. У другому розділі докладно описані методологія, дані та програмне забезпечення, які застосовувалися в цьому аналізі. У третьому розділі подаються ключові результати дослідження, а саме опис бізнес-моделей та міграція між ними. Також у третьому розділі показано ризиковість виявлених бізнес-моделей. Висновки стисло підсумовують результати дослідження та окреслюють перспективні напрями розробки тематики.

I. Огляд літератури

Розвитку досліджень бізнес-моделей банків за допомогою кластерного аналізу сприяли Базельські регуляції та впровадження процесу SREP. Дослідження бізнес-моделей страховиків є також перспективним напрямом досліджень.

Більша частина робіт направлена на аналіз та сегментацію клієнтів страхових компаній. Дослідження Сяочжи Вана та Імон Кіо (2008), Джессіки Заку (2019) присвячені кластерному аналізу для визначення цільової аудиторії компанії. Методи кластеризації використовувалися для визначення портрету клієнтів на основі наборів даних, отриманих з транзакцій полісів і профілів власників полісів. Робота реалізовувалась за допомогою карт самоорганізації та алгоритму К-середніх. Метод К-середніх також використано у публікації Шагаєга Аболмакарема та ін. (2016) направленої на сегментацію з метою визначення найбільш прибуткових клієнтів для компанії.

Більшість науковців використовують поєднання 2 методів кластеризації. Зокрема, у дослідженні проведеному Томиславом Павичем Крамаричем та іншими (2018) європейські страхові компанії розподілені на 7 кластерів. На відміну від попередніх досліджень тут кластеризуються компанії, а не їхні клієнти. За допомогою поєднання ієрархічної кластеризації та кластеризації методом К-середніх, 119 страховиків поділяються на 7 груп залежно від країни походження та типу компанії. Бах та інші (2020) використовують карту Кохонена у поєднанні з ієрархічним кластерним аналізом для дослідження ризику шахрайства лізингової компанії. Нейрони карти Кохонена поєднують у 5 кластерів за допомогою метода Уорда, після чого аналізують ризик-характеристики кластерів.

Робота Ансарі Салех Ахмар та інших (2018) є прикладом кластерного аналізу поза сферою фінансів. У своєму дослідженні вони за допомогою методу К-середніх групують провінції Індонезії. Таке групування, за думкою авторів, має допомогти класифікувати райони країни для легшого подолання соціальних проблем, що виникають. Аббас та інші (2020) порівняли методи К-середніх та К-медоїд, використавши дані про жінок у період вагітності. Метод К-медоїд є за своєю суттю дуже схожим з методом К-середніх, тому часто застосовується у поєднанні з ним. Автори показують, що для конкретних даних метод К-медоїд є більш точним ніж К-середніх.

Рашкован В. та Покідін Д. (2016) здійснили ідентифікацію бізнес-моделей банків в Україні за допомогою карти самоорганізації Кохонена та провели паралелі між бізнес-моделями та показниками різних типів ризику, на які банк може наражатись. За методологією наша робота є найбільше подібною саме до цього дослідження. Проте, на відміну від Рашкована та Покідіна, що базують висновки свого дослідження на карті самоорганізації Кохонена, ми використовуємо цей метод як додатковий до методу К-середніх. Також кластерний аналіз з метою виділення бізнес-моделей використали Ферстл Р. та Серес Д. (2012). На відміну від попередніх дослідників, вони використали алгоритм К-середніх, що базується на використанні відстані Махаланобіса. У роботі вдалось виділити 5 бізнес-моделей банків на основі 5 їхніх показників.

Більшість авторів використовують найпростіші моделі кластеризації, серед яких метод К-середніх. Ми у своїй роботі спробували розробити методологію, яка допомагає визначити розподіл компаній відповідно до їхньої бізнес-моделі. Для реалізації дослідження ми використали широкий набір інструментів клатеризації, проте висновки базували на методі К-середніх. Самоорганізаційні карти Кохонена у нашій роботі є зручним інструментом візуалізації. У даній роботі оцінка якості кластеризації вперше базується на використанні коефіцієнта міграції – показника, що характеризує стабільність кластерів.

II. ДАНІ ТА МЕТОДОЛОГІЯ

Опис використаних даних

Для проведення кластерного аналізу на основі звітності страхових компаній було зібрано дані, що включали річну звітність 247 українських страховиків за 2 роки, з 2019 по 2020 роки. За аналізований період кількість діючих страховиків суттєво зменшилась. Так на кінець 2020 року база даних складалась зі спостережень щодо 185 СТРАХОВИКІВ. Дані було зібрано з бази даних звітів, що страховики подають регулятору.

Для визначення бізнес-моделі ми намагалися відібрати показники, які б відповідали на такі питання щодо компанії:

- На яких клієнтів орієнтується?
- На яких видах страхування концентрується та наскільки явно?
- Які канали продажу використовує?

Під час аналізу було відібрано показники які водночас допомагали би дізнатись відповіді на ці питання та за допомогою яких розподіл на кластери відбувався б оптимально з огляду на кількісні метрики - коефіцієнт міграції та коефіцієнт силуету. Відповідно до отриманих значень був обраний оптимальний набір показників для кластеризації. Окремо було виділено показники, які дозволяють ширше описати кластери та ризики, що їм притаманні. Вони не увійшли в модель, оскільки розділення на їхній основі приводило до гірших показників якості кластеризації, але були використані для ширшої характеристики визначених кластерів. Таблиця 1 та Таблиця 2 описують алгоритм розрахунку цих характеристик.

Таблиця 1. Показники, за якими відбувався поділ на кластери.

№	Назва показника	Назва змінної	Формула
1	Рентабельність активів	ROA	Чистий прибуток / Вартість активів
2	Кількість представництв у регіонах	Offices	Сума офісів, що використовуються та не є головним офісом
3	Частка премій з обов'язкових видів страхування в загальному обсязі зібраних премій	% mandatory	Обсяг премій з обов'язкових видів страхування / Загальний обсяг премій
4	Частка премій від юридичних осіб	Corporate	Обсяг премій від юридичних осіб / Загальний обсяг премій
5	Частка вхідного перестраховання у преміях	Re-to-premiums	Обсяг премій з перестраховання / Загальний обсяг премій

Таблиця 2. Інформативні показники, які краще допомагають описати отримані бізнес-моделі.

№	Назва показника	Назва змінної	Формула
1	Співвідношення частки перестраховиків у страхових резервах до загального обсягу страхових резервів	Re-to-provisions	Сума прав вимоги до перестраховиків / Сума страхових резервів
2	Показник збитковості	Loss ratio	(Страхові Виплати + Витрати, пов'язані з регулюванням страхових випадків + зміна у резервах збитків / (Премії + зміна резерву незароблених премій)
3	Середній розмір зібраної премії	Mean premium	Сума зібраних премій / Кількість договорів страхування
4	Співвідношення заробітної плати до зібраних премій	Wages/Premiums	Сума витрат на заробітну плату / Сума зібраних премій за звітний період
5	Максимальний рівень концентрації на групі видів страхування	Concentration	Максимальне значення серед 7 категорій* / Загальний обсяг премій

* Перелік категорій: 1. Ядерне страхування; 2. Транспортне інше; 3. Автомобільне; 4. Відповідальності; 5. Особисте; 6. Майнове; 7. Інше.

Після розрахунку показників, їхні значення були стандартизовані (середнє значення приведені до 0, дисперсія – до 1). Це є необхідним з огляду на чутливість алгоритмів кластеризації до дисперсії у даних. Також у даних спостерігались викиди. Спостереження, що знаходяться далі трьох стандартних відхилень від середнього, було приведено до

найближчого значення, що входить у діапазон трьох стандартних відхилень. Розподіл спостережень до та після цієї процедури наводиться у Додатку 1.

Компанії, загальний обсяг премій яких за звітний період не перевищував 5 млн грн було віднесено до штучного кластеру (групи) – “Неактивні” у звітному періоді. Загальна частка на ринку (у преміях) таких компаній у 2020 р. складала менше 1%.

Нижче наводиться описова статистика даних за 2020 р.

Таблиця 3. Описова статистика змінних моделі.

	ROA	Offices	Re-to-premiums	% mandatory premiums	Corporate
Середнє	0.02	8.16	0.08	0.19	0.51
Ст. відхилення	0.10	24.99	0.21	0.26	0.33
Мін	-0.67	0.00	0.00	0.00	0.00
Q(25%)	0.00	0.00	0.00	0.00	0.22
Q(50%)	0.01	0.00	0.00	0.03	0.46
Q(75%)	0.05	0.00	0.03	0.35	0.86
Макс	0.40	200.00	1.00	0.86	1.00

Таблиця 4. Описова статистика інформативних показників компаній.

	Re-to-provisions	Loss ratio	Mean premium	Wages/Premiums	Concentration
Середнє	0.21	0.39	116.41	0.06	0.68
Ст. відхилення	0.24	0.51	405.89	0.06	0.19
Мін	0.00	-0.80	0.00	0.00	0.29
Q(25%)	0.02	0.11	0.89	0.02	0.54
Q(50%)	0.12	0.35	3.22	0.04	0.66
Q(75%)	0.36	0.52	29.79	0.07	0.83
Макс	0.93	4.08	3534.63	0.46	1.00

Можна побачити, що більшість компаній у 2020 році були малоприбутковими або збитковими, чого не спостерігалось у попередні роки (Додаток 2). Причиною такого падіння прибутковості можна було б назвати збільшення рівня виплат у страховиків, що займаються медичним страхуванням у зв'язку з пандемією COVID-19. Проте, проаналізувавши детальніше динаміку показників з медичного страхування, ми не побачили стрімкого зростання рівня виплат.

Більша частина ринку концентрує свою діяльність на дешевих договорах, які найімовірніше продають фізичним особам. Медіанний «середній чек» – близько 3 тис. грн.

За нашими даними, більшість компаній не мали офісів. З одного боку, це може свідчити про переважання на ринку бізнес-моделей, які не використовують офіси як канал продажу. З іншого боку, таке сильне зміщення вказує на можливий ризик неправильного звітування. Перевірити цю гіпотезу ми не можемо. Ми припускаємо, що компанії, які неправильно звітують, розподілені рівномірно між кластерами та не зміщують значно центри кластерів. Варто зазначити, що за даними 2019 року спостерігається аналогічна структура.

Також на ринку спостерігається висока концентрація на одній з груп видів страхування. Близько половини компаній мала частку премій з однієї з груп видів страхування більше 60%. Це свідчить про наявність спеціалізації компаній на певному сегменті страхування. З іншого боку, портфелі багатьох компаній можна назвати слабко диверсифікованими.

Методологія дослідження та моделі

Алгоритми кластеризації є зручним інструментом розділення спостережень на однорідні групи на основі заданих характеристик. Опис літератури наводить лише деякі з успішних випадків використання кластерного аналізу у дослідженні соціально-економічних явищ. Важливою перевагою таких алгоритмів є нівелювання впливу судження дослідника про досліджуване явище на результати дослідження. Для дослідження бізнес-моделей страховиків було використано класичні ієрархічні, неієрархічні методи кластеризації, та методи машинного навчання. Далі наводиться короткий опис застосованих методів.

Метод К-середніх є найбільш поширеним неієрархічним методом. Він полягає у ітеративній мінімізації відстані між спостереженнями у кластері, при цьому кількість кластерів задається на початку, тобто модель не визначає оптимальну кількість кластерів. Координати центроїдів, кількість яких відповідає кількості кластерів, на початковому етапі задаються випадково. Як результат, поділ на кластери може бути нестабільним та залежати від початкових координат центроїдів.

Для запобігання цій проблемі було використано метод початкових координат центроїдів "kmeans++", запропонований Девідом Артуром та Сергієм Васильєвським (2006). Позначивши вибірку вхідних даних χ , а найкоротшу відстань між елементом вибірки x_i та найближчим до нього центроїдом $D(x_i)$, алгоритм можна описати покроково:

1. Визначити випадково з χ початковий центроїд c_1 .
2. Наступний центроїд c_j визначити з χ , де кожному елементу присвоєно імовірність бути обраним $p(c_j = x_i) = \frac{D(x_i)^2}{\sum_{x \in \chi} D(x)^2}$
3. Повторювати крок 2 поки не буде визначено потрібну кількість центроїдів.
4. Застосувати класичний алгоритм К-середніх.

З кроку 2 видно, що елементам χ , що розташовані далі від початкового центроїду присвоюється більша імовірність бути обраним. Тобто центроїди розташовуються так, щоб бути відмінними одна від одної. Кластери на основі процедури "kmeans+" було оцінено 100 разів, з яких було обрано кластеризацію з найменшою сумою квадратів відстаней спостережень до центроїди (WCSS, within cluster sum of squares).

Підсумовуючи, центри кластерів спочатку обираються з елементів вибірки так, щоб розташовуватись далі один від одного, далі ітеративно змінюють свої координати щоб описувати найбільшу можливу групу (кластер) елементів вибірки.

Для методу K-середніх було обрано використати 5 кластерів, що є оптимальною кількістю кластерів з огляду на метод ліктя. Результати його застосування наведено у Додатку 3. Далі для усіх методів було використано розділення на 5 кластерів.

Метод K-медоїд, вперше описаний Кауфманом та Руссеу (1990) є за своєю суттю дуже схожим до методу K-середніх. Ключовою характеристикою методу K-медоїд є обрання центрів Кластерів серед спостережень. Такі спостереження, які є «еталонними» для свого кластеру, і називаються медоїдами.

За допомогою алгоритму PAM (partitioning around medoids, розділення навколо медоїд) медоїди обираються ітеративно таким чином, щоб зменшити середню відстань між спостережень до центрів їхніх кластерів. Сучасні алогоритми методу K-медоїд пропонують швидшу оптимізацію, проте PAM залишається одним із найточніших алгоритмів вирішення цієї задачі. Тому нами було обрано саме це метод. Початкові медоїди нами було обрано за допомогою алгоритму "kmedoids+", що є ідентичним до "kmeans+" та забезпечує стійкість кластерів.

Ієрархічні методи не потребують визначення кількості кластерів перед застосуванням алгоритму. Вони будують деревовидну структуру, що називається дендрограмою. На початку кожне спостереження формує окремий кластер. Далі спостереження (кластери) на основі обраного критерію поєднуються у нові, більші кластери, поки не з'єднаються в один кластер, в який входять усі спостереження. Для нашої цілі було обрано **критерій Уорда**. Кількість кластерів визначається дослідником на основі дендрограми вже після застосування алгоритму.

За критерієм Уорда, окремий кластер поєднується з тим кластером, поєднання з яким призведе до найменшого зростання відстані між спостереженнями усередині кластеру. Ця відстань, яка є аналогічною до WCSS алгоритму K-середніх, зображається на дендрограмі по вертикальній осі.

Далі наводиться короткий опис **карти Кохонена**, що є методом машинного навчання, здатним до кластеризації. Його описано у роботі Тойво Кохонена (1990). Карта самоорганізації є штучною нейронною мережею, що складається з двох шарів:

1. Вхідного, на який подаються дані вибірки. Розмірність цього шару відповідає кількості характеристик, за якими відбувається поділ на групи.
2. Вихідного, що являю собою власне карту, що складається з впорядкованих у двох (у випадку даної роботи) вимірах нейронів, і має наперед визначену довільну розмірність.

Кожен нейрон карти поєднаний з кожним елементом вхідного шару вузлами, що мають ваги. Тобто кожен нейрон має набір ваг, що можуть трактуватись як опис нейрону у характеристиках, що відповідають даним вхідного шару. Алгоритм навчання карти Кохонена можна описати покроково:

1. Ваги нейронів на початку навчання є випадковими достатньо малими числами.
2. На вхідний шар подається вектор ознак x_i з χ та розраховується відстань (у цьому дослідженні використано квадрат евклідової відстані) між векторами x_i та w_j , де w_j – вектор ваг нейрону j вихідного шару мережі.
3. Нейрон, що є найближчим до x_i на основі кроку 2, іменується нейроном-переможцем (BMU, best matching unit).
4. Виходячи з радіусу $\sigma(t)$, для кожного нейрону карти розраховується показник сусідства на основі функції Гауса $N(t)_{BMU,j} = \exp\left(-\frac{D(BMU,j)^2}{2\sigma(t)^2}\right)$, де $D(BMU,j)$ – топографічна відстань між нейроном-переможцем та нейроном j .
5. Ваги нейронів на карті оновлюються за формулою $\Delta w_j = \alpha(t)N_j(t)(x_i - w_j)$, де $\alpha(t)$ – коефіцієнт навчання, що є спадною функцією від часу.

6. Кроки 2-5 повторюються для заданої кількості епох (циклів навчання), що визначається дослідником. При цьому прийнято звертати увагу на похибку квантування, що відображає середню відстань між вхідними даними та нейроном-переможцем, та топографічну похибку, що відображає кількість спостережень, для яких перший нейрон-переможець не є сусідом другого нейрона переможця.

У результаті навчання нейрони стають «схожими» на вхідні дані. У процесі навчання параметри $\alpha(t)$ та $\sigma(t)$ поступово зменшуються. Таким чином, чим далі заходить навчання, тим повільніше нейрони адаптують свої ваги та тим меншу «кооперацію» демонструють. Використана у цій роботі спадна функція, що описує динаміку параметру $\alpha(t)$, має вигляд:

$$\alpha(t) = \frac{\alpha(0)}{1+(t/(MI/4))} \quad (1)$$

де $\alpha(0)$ – початкове значення $\alpha(t)$, що задається дослідником;

MI – максимальна кількість епох (ітерацій), що задається дослідником;

t – порядковий номер епохи.

$\alpha(0)$ нами для цього дослідження було обрано на рівні 0.5, MI – 10000. Таким чином, коефіцієнт навчання поступово затухає від 0.5 до 0.1. Динаміка параметру $\sigma(t)$ у процесі навчання моделі є аналогічною, з початковим значенням 1. Більш детально алгоритм самоорганізаційної карти пояснено у роботі Хонкела (1998).

Для цього дослідження, з огляду на кількість спостережень, було обрано використати карту розмірністю 10x10 (100 нейронів загалом) прямокутної топології. Поширеною є ініціалізація ваг нейронів на основі головних компонент, що спостерігаються в даних. Проте, з огляду на те, що нейрони активуються (є переможцями) рівномірно на карті (Додаток 4), та швидкість навчання є прийнятною, ми не стали застосовувати цей метод. Динаміку топографічної помилки та помилки квантизації представлено у Додатку 5.

Після навчання для можливості співвідношення результатів карти Кохонена з іншими методами нейрони було згруповано, застосувавши до їхніх ваг метод K-середніх.

Описані методи було реалізовано інструментами Python за допомогою відкритих бібліотек sklearn (Pedregosa et al., 2011) та MiniSom (Vettigli, 2020).

Оцінка результатів кластеризації.

Кожен з методів має свої переваги та недоліки, їхню оцінку наведено у Таблиці 5.

Таблиця 5. Порівняння методів кластеризації.

Характеристика	К-середніх	К-медоїд	Метод Уорда	Карти Кохонена
Легкість інтерпретації результатів	+	+	+	-
Наявність графічних інструментів	-	-	+	+
Стійкість до викидів	-	+-	+-	+-
Можливість застосувати оцінену модель до різних наборів даних	+	+	-	+
Легкість у побудові	+	+	+	-

Як видно, жоден з методів не можна визначити однозначно як найкращий. Тому найчастіше при застосуванні кластерного аналізу обирають метод, що найкраще підходить з огляду на наявні дані та чисельні критерії якості поділу на кластери.

Для оцінки якості моделей ми використали класичний показник **CH**-score, який оцінює якість розділення на групи на основі відстаней між спостереженнями. Також, для наших цілей важливим є стійкість кластерів у часі. Бізнес-моделі відображають стійку поведінку (стратегію), а отже для того, щоб робити висновки про бізнес-моделі та їхні ризики, важливо щоб кластери не змінювались суттєво у часі. Для оцінки стійкості ми застосували коефіцієнт міграції.

CH-score (Calinski-Narabasz score) вперше описано Калінським та Харабазом (1974). Показник відображає відношення варіації між кластерами до варіації всередині кластерів, обидва зважені на відповідні ступені свободи. Показник обчислюється за формулою:

$$CH - score = \frac{\left(\frac{BCSS}{WCSS} \right)}{n - k} \quad (2)$$

де $BCSS$ – варіація між кластерами;

$WCSS$ – варіація між спостереженнями всередині кластерів;

k – кількість кластерів;

n – кількість спостережень.

Еталонного значення цього показника не існує. Проте більше його значення свідчить про краще розділення на кластери. Значення CH -score більшим тоді, коли центри кластерів перебувають далі один від одного, а спостереження у кластерах є близькими до їхніх центрів.

Якісне розділення утворює групи (кластери), що не змінюються суттєво з часом. У нашому прикладі це є принципово важливим, оскільки бізнес-модель – це стала характеристика компанії, яка не змінюється суттєво за нормальних умов функціонування. Для оцінки стабільності кластерів було розраховано коефіцієнт міграції:

$$\text{Коефіцієнт міграції} = \frac{n_m}{n_{2019 \cap 2020}} \quad (3)$$

де n_m – кількість компаній, що на основі моделі мігрували між кластерами з 2019 по 2020 р.

$n_{2019 \cap 2020}$ – кількість компаній, що були активними у 2019 та 2020 роки.

Міграція між кластерами відбувається у результаті дії двох факторів – зміни бізнес-моделі компанії та похибки кластеризації. Тому завелике значення коефіцієнта міграції свідчить про неточне розділення, а замале – про «перетренованість» моделі. Критичного значення цього показника немає.

Для методу Уорда було розраховано псевдо коефіцієнт міграції. Оскільки оцінену модель не можна застосувати до даних за інший рік, ми розраховали ПСЕВДО коефіцієнт міграції. Для розрахунку псевдо коефіцієнта міграції було оцінено модель на найактуальніших даних. Далі на основі класифікаційного методу на основі центроїдів, описаного Тібширані та іншими (2002) було визначено кластери для даних попереднього року та застосовано формулу (3).

У Таблиці 6 наведено оцінки якості кластеризації для застосованих моделей.

Таблиця 6. Порівняння якості методів кластеризації.

Показник	К-середніх	К-медоїд	Метод Уорда	Карта Кохонена
CH-score	68.807	68.806	77.101	-
Коефіцієнт міграції	15.8%	19.0%	20.6% (псевдо)	15.8% (між кластерами), 76.9% (між нейронами)

Видно, що метод Уорда з огляду на критерій CH-score дає найкращі результати кластеризації. Проте, кластери є значно менш стабільними у порівнянні з усіма іншими МЕТОДАМИ. З огляду на цей критерій, який має велике значення з точки зору застосовуваності моделі, нами було прийнято рішення не робити висновки на основі методу Уорда. Довідково результати методу Уорда представлено у Додатку 6. Для зручності порівняння, кластери було проіменовано аналогічно до основних моделі.

Моделі К-середніх та К-медоїд мають дуже близькі значення CH-score. Також ці значення, хоч і є нижчими за отриманий при застосуванні методу Уорда, проте не надто суттєво. Модель К-середніх показує більш стабільні кластери, ніж модель К-МЕДОЇД. Також, за наявності зміщень у даних, модель К-медоїд може не характеризувати повністю кластери, оскільки базує висновки на одному спостереженні. Наприклад, чотири з п'яти кластерів, ця модель характеризує як бізнес-моделі, що взагалі не використовують представництва у своїй діяльності. Такий висновок є помилковим з огляду на результати моделі К-середніх, що представлені далі. Тому, зважаючи на викривлення даних, ми не базуємо свої висновки на результатах моделі К-медоїд. Довідково результати методу К-медоїд представлено у Додатку 7.

Для карти Кохонена було розраховано два види коефіцієнта міграції. Перший – на основі п'яти кластерів, у які згруповано нейрони. Другий – на основі усіх ста нейронів моделі. Очікувано, перший є значно меншим за другий. Цікаво, що міграція між кластерами при застосуванні карти Кохонена є майже ідентичною до методу К-середніх.

З огляду на результати оцінки якості моделей, ми вирішили базувати свої висновки на результатах моделі К-середніх. Також, оскільки результати карти Кохонена є схожими з результатами К-середніх, ми використали її як інструмент візуалізації кластерів.

Далі, ми побудували графік силуету (Rousseeuw P., 1987) для результатів методу К-середніх, щоб більш детально оцінити результати. Коефіцієнт силуету розраховується для кожного спостереження як:

$$s_i = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4)$$

де $a(i)$ – середня відстань від спостереження i до членів свого кластеру;

$b(i)$ – середня відстань від спостереження i до членів найближчого сусіднього кластеру.

Значення коефіцієнту для моделі є середнім значенням коефіцієнту силуету усіх спостережень. Значення коефіцієнту силуету 1 свідчить про ідеальне розмежування кластерів, 0 – розмежування є випадковим, (-1) – кластери розмежовано неправильно.

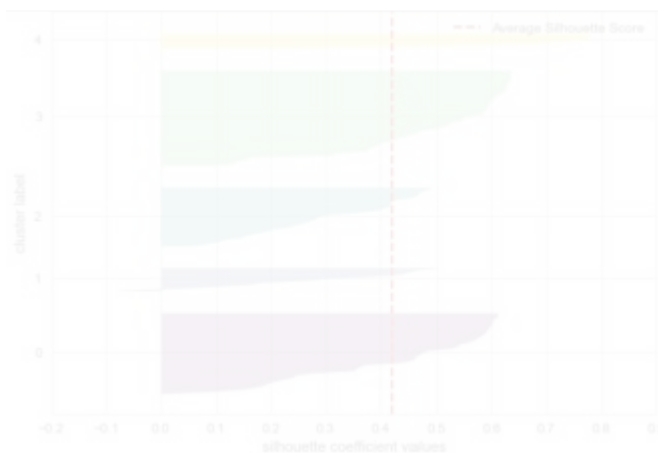


Рисунок 1. Силуети кластерів (абсциса – коефіцієнт силуету, ордината – номер кластеру)

Графік коефіцієнтів силуету для спостережень свідчить про те, що лише для одного спостереження сусідній кластер є «ближчим» ніж свій. Це спостереження, та ті що мають значення силуету близьке до 0 можуть лежати «на межі» кластеру. На Рисунок 1, за від'ємним значенням, видно, що дійсно у кластері 1 – Універсальна “великі” наявна компанія, що є віддаленою від центру кластера, але і не є близькою до іншого. Загальне значення коефіцієнту (0.41) свідчить про достатньо якісне розділення груп, при цьому з графіка видно, що кластер 4 “Перестрахова” модель ідентифіковано найкраще.

III. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Опис бізнес-моделей на основі кластерів

Модель було оцінено на даних 2020 року та застосовано до усіх років, що наявні у вибірці (2019-2020 рр.). Характеристики (координати центроїд) ідентифікованих кластерів зображено на Рисунку 2. Координати на рисунку є стандартизованими.



Рисунок 2. Характеристики ідентифікованих кластерів (стандартизовано).

Ідентифіковані кластери для зручності було пронумеровано та проіменовано. Ми не виключаємо, що інші однорідні групи компаній можуть бути ідентифіковані з даних та можуть викривлювати характеристики ідентифікованих нами кластерів, проте спираючись лише на дані та методи, описані вище, нам вдалось ідентифікувати ті бізнес-моделі, які найкраще виділяються з огляду на кількісні критерії. Далі, на основі характеристик, за якими було здійснено розподіл, та описових ПОКАЗНИКІВ, наводиться короткий огляд ідентифікованих кластерів (бізнес-моделей), розрахованих для кожного кластеру. Додаток 8 підсумовує результати кластеризації у нестандартизованому вигляді.

“0 – Роздрібна” бізнес-модель концентрується на страхуванні фізичних осіб (74% фізичних осіб у преміях) та мають середній рівень рентабельності активів (3%). З огляду на невелику кількість представництв (офісів), значне відношення заробітної плати до премій, компанії цієї бізнес-моделі як канал продажу використовують власних агентів, які намагаються знайти клієнтів. Тобто, компанії «полюють» на клієнтів, а не клієнти самі до них звертаються. Страховики, що використовують цю бізнес-модель спеціалізуються на страхуванні добровільних видів, та мають частку перестраховання у преміях у середньому на рівні 1.5%. Варто зазначити, що даний кластер має найбільший показник концентрації (76%) на одній з груп видів страхування, що може бути фактором ризику компаній цієї групи. На 2020 р. у компаній кластеру спостерігався найбільший показник збитковості, та помірна частка перестраховання у резервах, що свідчить про вразливість компаній до андеррайтингового ризику. До цього кластеру у 2020 році увійшло 40 компаній, що становить 19% ринку за преміями.

Модель “1 – Універсальна “великі”” – обслуговує як юридичних, так і фізичних осіб, має розподіл між обов’язковим та добровільним страхуванням 28% та 72% відповідно. Рентабельність активів є найвищою серед усіх виділених кластерів (6.5%). Характерною ознакою кластеру є широке використання власних офісів (мають в середньому близько 62 представництва). Ознакою того, що компанії цієї бізнес-моделі активно використовують, як власні представництва так і агентів у якості каналу продажу, є висока частка заробітних плат у преміях (6%). Отже, компанії намагаються диверсифікувати способи залучення клієнтів. Також компанії цього кластеру мають другу найбільшу частку ризику, що перестраховується (25%), що свідчить про менші андеррайтингові ризики, а також про те, що компаніям властивий ризик дефолту контрагента (перестраховика). 12 компаній у 2020 році використовували цю бізнес-модель і займали найбільшу частку за валовими преміями – 35.5%. Варто зазначити, що з огляду на високу частку ринку у поєднанні з найнижчим показником середньої премії, компанії орієнтуються на продаж великої кількості дешевих полісів.

“2 – Універсальна “малі”” описується порівняно рівномірним розподілом обов’язкового і добровільного страхування та фізичних і юридичних осіб у преміях (64%/36% та 63%/37% відповідно). При цьому частка премій з обов’язкових видів страхування у цій бізнес-моделі є найбільшою серед усіх кластерів. Додатковий аналіз показав, що премії з ОСЦПВ становлять

71% премій з обов'язкових видів для даного кластеру. Також кластер має другий найнижчий серед усіх груп показник рентабельності активів, компанії володіють в середньому 6 представництвами (офісами) та мають другу найвищу частку перестраховання у преміях (5.6%). Окрім орієнтації на обов'язкові види, дана модель відрізняється від Універсальна "великі" значним показником середньої премії, що вказує на те, що компанії намагаються страхувати більш дорогі ризики. До цього кластеру у 2020 році увійшли 29 компаній, які сукупно складають близько 16.5% ринку за преміями.

"3 – Корпоративна" модель характеризується 89% часткою юридичних осіб у преміях, а також невисоким показником рентабельності активів (2.7%), маленькою кількістю представництв (офісів) та близькою до 0 часткою обов'язкового страхування. Виділяється значна, як для компаній що не використовують перестрахову бізнес-модель, частка вхідного перестраховання у преміях (27%). При порівняно високому рівні середньої премії (254 тис. грн) компанії цього кластеру мають досить низький показник збитковості серед усіх ідентифікованих бізнес-моделей (22.5%). Передбачуваною є висока частка перестраховика у резервах, адже часто таким страховикам необхідно розділити велику експозицію корпоративного клієнта. Проте, це створює компаніям кластеру ризик дефолту контрагента (перестраховика). У цей кластер увійшла найбільша кількість компаній – 47, що за преміями 2020 року сукупно становить 19% ринку.

Компанії бізнес-моделі "4 – Перестрахова" мають частку перестраховання у преміях в середньому близько 81%. Рентабельність активів у компаній даного кластеру у середньому є від'ємною, середня премія – більше 304 тис. грн. Компанії цього кластеру взагалі не використовують представництва як канал продажу та мають найнижчу частку заробітних плат у преміях. Частка добровільного страхування у преміях наближається до 100%. Перестраховики самі є слабо перестрахованими, що може свідчити про потенційну вразливість до андеррайтингового ризику, який вони не розділяють (не диверсифікують) між собою. Проте, низьке у порівнянні з іншими бізнес-моделями значення показника збитковості свідчить про незначну реалізацію андеррайтингового ризику. У кластер увійшли 8 компаній за даними 2020 року (10% ринку за преміями).

Гістограми ознак для отриманих кластерів зображено у Додатку 9.

Нейрони на самоорганізаційній карті Кохонена у процесі тренування починають своїми вагами бути “схожими” на вхідні дані, тобто відтворювати кластери. Карта самоорганізації Кохонена представляє зручний інструмент візуалізації подібності між різними спостереженнями та характеристик цих спостережень. За допомогою карти можна побачити спостереження, що лежать на межі кластерів, та наскільки вони віддалені від інших елементів кластера.

На картах ознак на Рисунку 3 зображено ваги нейронів карти, що відповідають ознакам даних (координати стандартизовані). Їх варто інтерпретувати так: компанії, для яких нейрон з координатами (1;10) (лівий верхній кут) після навчання є переможцем (найближчим нейроном), мають середню по ринку частку премій з обов’язкових видів страхування, багато представництв та майже відсутнє перестраховання у преміях.

Видно, що сусідні нейрони є доволі схожими через механізм “кооперації” під час навчання, тож їх можна об’єднати у групи. Для цього використано алгоритм K-середніх з кількістю кластерів, що дорівнює 5, щоб інтерпретувати результати схожим чином як результати поділу методом K-середніх. Ми не наводимо центроїди цих кластерів, оскільки групування методом K-середніх було здійснено лише для того, щоб математично оцінити межі кластерів на карті Кохонена, та такі центроїди повторювали би за своєю суттю центроїди попередньої моделі.





Рисунок 1. Карти характеристик Кохонена.

На Рисунок 4 наведено результати об'єднання нейронів у кластери. Точки на рисунку позначають компанії, для яких конкретний нейрон після навчання є нейроном-переможцем. Як видно, з огляду на об'єднання нейронів та карти ознак можна виділити бізнес-моделі, що є аналогічними ідентифікованим методом К-середніх.

Проте, карта дозволяє побачити відстань (подібність) між спостереженнями. Топографічну відстань можна побачити відразу на карті – сусідні нейрони є схожими. Евклідову відстань між нейронами після навчання можна побачити у Додатку 10. Компанії, для яких нейрон-переможець розташований на топографічній межі кластеру, є невзірцевими представниками кластеру та можуть змінювати свій кластер з часом. Саме про такі компанії йдеться у частині валідації результатів моделі К-середніх.



Рисунок 2. Групування кластерів на карті Кохонена

Коефіцієнт міграції для карти Кохонена між п'ятьма кластерами складає 5%, між 100 нейронами – 80%. Це свідчить про те, що міграція всередині кластерів є більшою, ніж між кластерами. Проте, з огляду на цей показник, висновки щодо бізнес-моделей наводяться виходячи з результатів кластеризації методом К-середніх. Але варто помітити, що економічна суть моделей, визначених за допомогою карти Кохонена, збігається з результатами методу К-середніх.

Аналіз міграції між кластерами

Маючи виділені кластери (бізнес-моделі), можна прослідкувати динаміку складу кластерів упродовж досліджуваного періоду. Було розраховано коефіцієнти міграції з кожного кластеру у період з 2019 по 2020 рр.

Як описано вище, міграція між кластерами може відбуватись під впливом двох факторів – похибки моделі та зміни компанією бізнес-моделі. Розуміючи, що кластеризація не є точним методом, ми вважали міграцію між кластерами значимою, якщо з кластеру мігрувало більше 10% компаній, що до нього входять. Застосування такого порогу показує значимими міграції щонайменше двох компаній з кластеру та здебільшого міграції, що є більшими за значення коефіцієнта міграції моделі (15%).

Оскільки компанії, що заробили за рік премій менше ніж 5 млн грн, не було включено до алгоритму K-середніх та віднесено до штучного кластеру “Неактивні”, існує міграція з виділених бізнес-моделей у штучний кластер “Неактивні”.

Рисунок 5 показує значимі міграції компаній між бізнес-моделями. Видно, що суттєва міграція у 2020 р. існувала з моделі Універсальна «великі» до моделі Універсальна «малі». Це можна вважати очікуваним, адже відмінність кластерами моделями не є значною. Менш очевидною є міграція компаній з моделі Перестрахова до моделі Корпоративна. Проте, якщо подивитись на центри кластерів, можна побачити що Корпоративна модель є ближчою до Перестрахової ніж інші, адже у компаній Перестрахової моделі є невелика частка премій від юридичних осіб, а у компаній моделі Корпоративна – частка премій з вхідного перестраховування.



Рисунок 3. Міграції між бізнес-моделями

З огляду на міграції до групи “Неактивні”, компанії, що використовують бізнес-моделі Корпоративна, Роздрібна та Перестрахова мають більший ризик виходу з ринку, а отже є менш стійкими. Більше половини компаній моделі Перестрахова вийшли з ринку у 2020 р., що може слугувати сигналом для регулятора щодо пильного нагляду.

Описуючи бізнес-моделі, ми зауважили, що Корпоративні страховики широко використовують вихідне перестраховання своїх ризиків, тобто у своїй роботі залежать від Перестраховиків. Отже, логічним є високий рівень виходу з ринку компаній саме цих двох кластерів разом.

Можна оцінити робастність висновків, що базуються на міграції до групи «Неактивні». У цю групу ми відносили компанії, обсяг премій яких за рік був меншим за 5 млн грн, отже мігрувати могли дуже малі компанії, для яких премії близько 5 млн грн є нормою, через варіацію обсягу премій рік від року. Менше 10 млн грн премій у 2019 р. мали лише 24% компаній, що мігрували у «Неактивні». Медіанне значення обсягу премій, серед компаній, що мігрували у групу становить 175 млн грн, середнє – 397 млн грн. Оскільки різке скорочення обсягу премій з таких значень свідчить про кризу у діяльності компанії, результати можна вважати робастними. Лише 16% цих компаній мали ненульовий обсяг премій у 2020. Тобто більшість повністю зупинила страхову діяльність.

Більш детально міграції між кластерами та їхні причини можна оцінити за допомогою карти Кохонена. На Рисунку 6 зображені компанії, які або у 2019 р. або у 2020 р. входили у відповідні кластери. Варто звернути увагу, що якщо Рисунок 5 показує односторонні міграції, то на карті Кохонена зображено двосторонні (разом компанії які мігрували з кластеру в кластер).

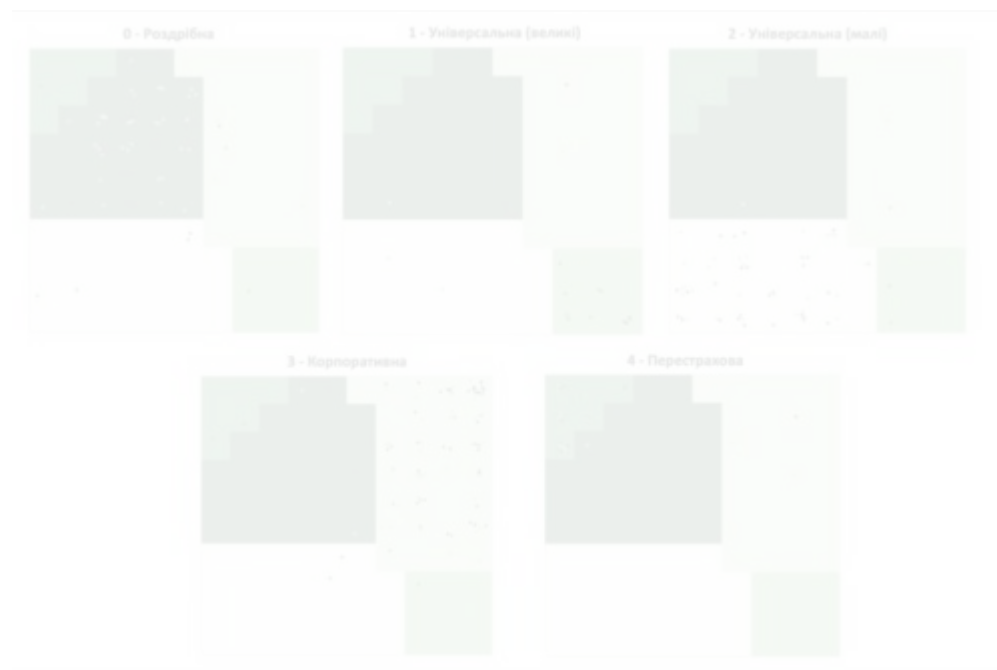


Рисунок 6. Міграції на карті Кохонена

На рисунку можна спостерігати, що міграція відбувається здебільшого між сусідніми кластерами та нейронами. Коефіцієнт міграції між кластерами для карти Кохонена є близьким до моделі K-середніх, він становить 15.8%.

Наприклад, видно що значна міграція з кластеру Перестрахова до кластеру Корпоративна, яку ідентифіковано за обома моделями відбувається за рахунок різкого згортання перестраховиками своєї перестрахової діяльності та початку обслуговування напряму корпоративних клієнтів. Оскільки ця міграція відбувається не між сусідніми нейронами та кластерами, можна стверджувати, що компанії не перебували на межі кластерів, а дійсно суттєво змінили бізнес-МОДЕЛЬ.

На відміну від моделі К-середніх, на карті Кохонена ми не спостерігаємо значної міграції всередині бізнес-моделі Універсальна. Обидва кластери цієї моделі демонструють незначну міграцію компаній до та з інших кластерів, яку ми не можемо назвати стійкою. Цікавим є те, що для малих компаній цієї бізнес-моделі міграція відбувається здебільшого з сусідніми нейронами на межі кластеру, для великих – міграція тільки далеко від межі кластеру.

Зобразивши компанії, що перестали займатись страховою діяльністю на карті Кохонена, можна виділити її зони, що характеризуються високим ризиком (Рисунок 7).

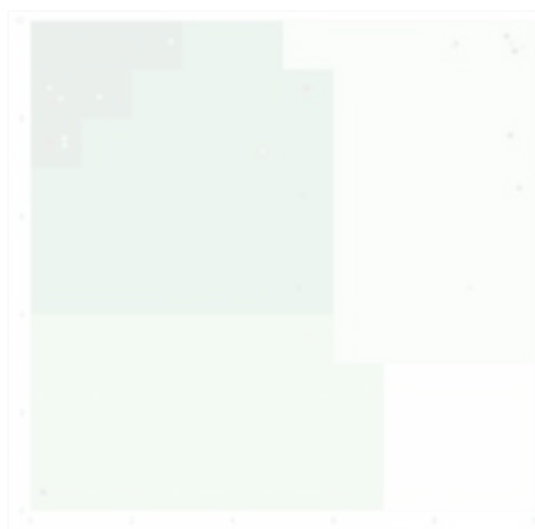


Рисунок 7. Компанії, що мігрували до групи «Неактивні».

Результати карти Кохонена є консистентними з моделлю К-середніх, найбезпечніша бізнес-модель – Універсальна. Компанії, що використовують моделі Перестрахова та Корпоративна емпірично є найменш стійкими.

Особливо ризиковою зоною моделі Корпоративна є правий верхній кут карти. Там знаходяться компанії, у яких частка юридичних осіб у преміях близька до 100% та які страхують добровільні види. Варто зазначити, що збитковість навряд є причиною ризиковості цих компаній, адже вони мають близький до середньоринкового рівень рентабельності активів.

Для моделі Перестрахова ризиковою є нижня частина кластеру на карті. Це компанії, які займаються як перестрахованням, так і прямим страхуванням добровільних видів. Можна зробити висновок, що більш стабільні перестраховики займаються лише перестраховою діяльністю.

Можна помітити, що регіон карти, що характеризується найвищою рентабельністю активів не показує абсолютно міграцій до групи Неактивні. Тобто, прибуткова діяльність підвищує стабільність компаній.

ВИСНОВКИ

Ефективний нагляд за страховим ринком будується на розумінні ключових ризиків, які притаманні діяльності компаній. Визначення бізнес-моделей у цьому процесі може відіграти важливу роль, оскільки бізнес-модель описує ключові характеристики бізнесу компанії, а отже і перспективні ризики та слабкі місця.

У роботі проведено кластерний аналіз, який базується на методі K-Середніх та нейронних мережах Кохонена, з метою ідентифікації бізнес-моделей страхових компаній. На основі кількісних даних щодо діяльності компаній було виділено чотири різних бізнес-моделі страховиків на ринку України: Роздрібна, Корпоративна, Універсальна (поділені на 2 кластери), Перестрахова. Шостий кластер є штучно виділеним, до нього увійшли страхові компанії, валові премії яких за рік склали менше 5 млн грн та які для цілей цього дослідження вважались неактивними. Також запропоновано опис виділених бізнес-моделей на основі ключових кількісних показників, що їх характеризують.

Компанії роздрібної бізнес-моделі страхують фізичних осіб, та схильні концентруватись на певній групі видів страхування. Така концентрація та низький рівень вихідного перестраховання робить їх вразливими до андерайтингового ризику.

Великі універсальні страховики – це здебільшого відомі страхові компанії, які користуються довірою споживачів, мають багато офісів та високу рентабельність. Вони орієнтуються на продаж великої кількості дешевих полісів.

Малі універсальні страховики схильні страхувати обов'язкові види, зокрема ОСЦПВ. Таким чином, ризики цієї бізнес-моделі тісно пов'язані з ризиками автоцивільної відповідальності. Ці компанії мають низьку рентабельність.

Корпоративні страховики концентруються на юридичних особах та страхують дорогі ризики. Вони широко використовують вихідне перестраховання для зменшення андерайтингового ризику. Проте, це робить їх вразливими до ризику дефолту контрагента.

Перестраховики є найменш рентабельними на ринку, перестраховують здебільшого добровільні види страхування. Ми виявили, що перестраховики самі є слабко перестрахованими, що робить їх вразливими до андерайтингового ризику.

Модель кластеризації, використана у нашій роботі, може допомогти у процесі нагляду. Визначення бізнес-моделі страховика на основі моделі зменшує витрати часу на аналіз фінансової інформації об'єкта нагляду. Важливим є також зменшення впливу Суб'єктивного судження у цій частині нагляду.

У дослідженні показано міграції страхових компаній між кластерами. Відповідно до моделі, компанії, що використовують Корпоративну та Перестрахову бізнес-модель з 2019 по 2020 рр. найчастіше виходили з ринку, що може свідчити про те, що такі компанії потребують більшої уваги регулятора. В той же час найбільш стабільними, а отже і найменш ризиковими є Роздрібна та Універсальна моделі. Отже, запропоноване поєднання методів можна вважати ефективним для цілей нагляду за ринком.

Ця робота дає підґрунтя для подальших досліджень у двох сферах. По-перше, ми вважаємо ідентифіковані у цій роботі кластери доволі широкими, хоча вони і відповідають ключовим напрямкам діяльності компаній. Отже, виділення бізнес-моделей, що є більш вузько орієнтованими, на основі ідентифікованих у цій праці кластерів було би логічним продовженням розробки теми. По-друге, з огляду на описану емпіричну залежність рівня ризику страховика від бізнес-моделі, яку він використовує, вкрай важливим є дослідження факторів ризику, що діють на компанії різних кластерів. Ключовим чинником, що сприяв би подальшій розробці даної тематики ми бачимо доступність детальних даних щодо компаній на страховому ринку України.

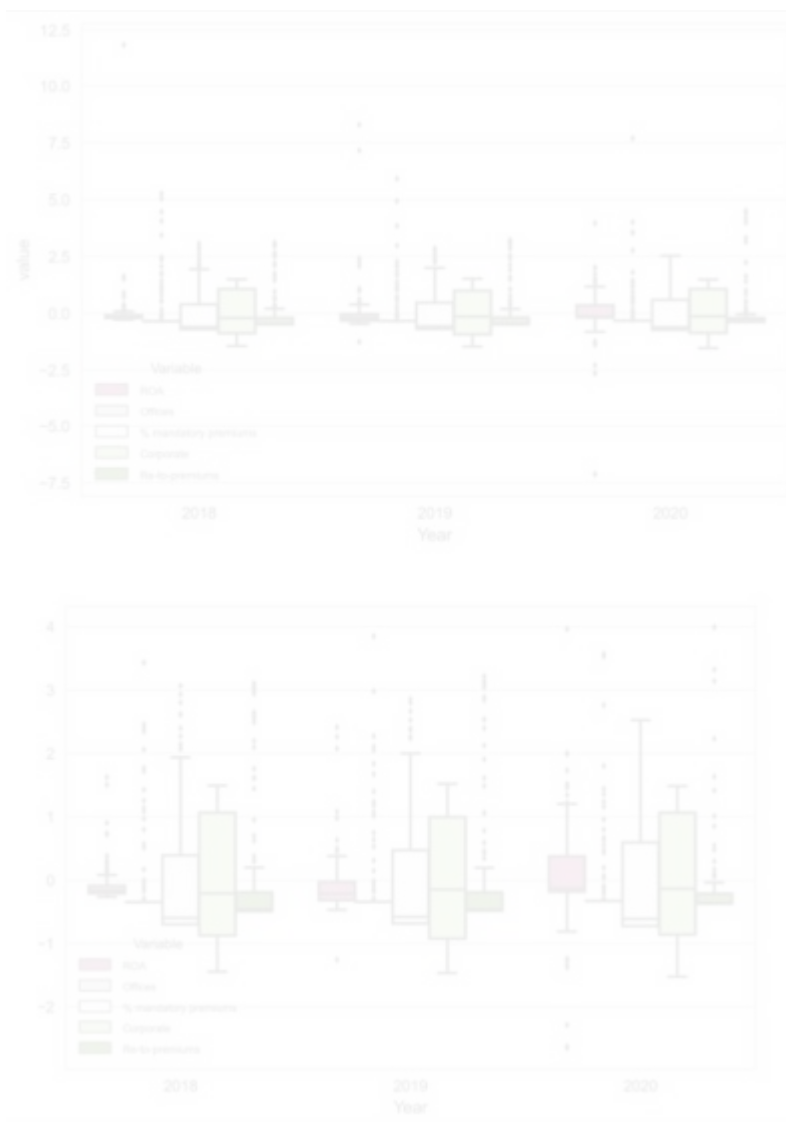
Список літератури

1. A.S., A., Napitupulu, D., Rahim, R., Hidayat, R., Sonatha, Y., & Azmi, M. (2018). Using K-Means Clustering to Cluster Provinces in Indonesia. *Journal of Physics: Conference Series. 2nd International Conference on Statistics, Makassar, Indonesia*. <https://doi.org/10.1088/1742-6596/1028/1/012006>
2. S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif and S. Z. H. Kazmi, "K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir," in *IEEE Access*, vol. 8, pp. 151847-151855, 2020, doi: 10.1109/ACCESS.2020.3014021.
3. Ansari Saleh Ahmar et al 2018 *J. Phys.: Conf. Ser.* 1028 012006 <https://doi.org/10.1088/1742-6596/1028/1/012006>
4. Arthur, D. & Vassilvitskii, S. (2007). *k-means++: the advantages of careful seeding*. SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (p./pp. 1027--1035), Philadelphia, PA, USA: Society for Industrial and Applied Mathematics. ISBN: 978-0-898716-24-5
5. Bach, M., Vlahović, N. & Pivar, J. (2020). *Fraud Prevention in the Leasing Industry Using the Kohonen Self-Organising Maps*. *Organizacija*, 53(2) 128-145. <https://doi.org/10.2478/orga-2020-0009>
6. Honkela, T. (1998, February 2). *Description of Kohonen's Self-Organizing Map*. Aalto University. Retrieved January 20, 2022, from <http://www.mlab.uiah.fi/~timo/som/thesis-som.html>
7. Kaufman, L. & Rousseeuw, P. J. (2008). *Partitioning Around Medoids (Program PAM). Finding Groups in Data* (pp. 68--125). John Wiley & Sons, Inc.. ISBN: 9780470316801
8. Kohonen, T. *Self-organized formation of topologically correct feature maps*. *Biol. Cybern.* 43, 59–69 (1982). <https://doi.org/10.1007/BF00337288>
9. Kramarić, T. P., Bach, M. P., Dumičić, K., Žmuk, B., & Žaja, M. M. (2017). *Exploratory study of insurance companies in selected post-transition countries: non-hierarchical cluster analysis*. *Central European Journal of Operations Research*, 26(3), 783–807. <https://doi.org/10.1007/s10100-017-0514-7>
10. Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others. (2011). *Scikit-learn: Machine learning in Python*. *Journal of Machine Learning Research*, 12(Oct), 2825–2830.

11. Peter J. Rousseeuw (1987). *Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis*. *Computational and Applied Mathematics*. 20: 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
12. Rashkovan, V., Pokidin, D. (2016). *Ukrainian Banks' Business Models Clustering: Application of Kohonen Neural Networks*. *Visnyk of the National Bank of Ukraine*, 238, 13-38.
13. Robert Ferstl & David Seres, 2012. "Clustering Austrian Banks' Business Models and Peer Groups in the European Banking Sector," *Financial Stability Report*, Oesterreichische Nationalbank (Austrian Central Bank), issue 24, pages 79-95.
14. Shaghayegh A., Farshid A., Kaveh K.-D. (2016). *Insurance customer segmentation using clustering approach*. *International Journal of Knowledge Engineering and Data Mining (IJKEDM)*, Vol. 4, No. 1, 18-39. DOI: 10.1504/IJKEDM.2016.082072
15. Tibshirani, R., Hastie, T., Narasimhan, B., & Chu, G. (2002). *Diagnosis of multiple cancer types by shrunken centroids of gene expression*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(10), 6567-6572. *The National Academy of Sciences*.
16. Vettigli, G. *MiniSom: Minimalistic and NumPy-Based Implementation of the Self Organizing Map*. Release 2.1.5. 2019 (accessed 12 August 2019); <https://github.com/JustGlowing/minisom>
17. Wang X., Keogh E. (2008) *A Clustering Analysis for Target Group Identification by Locality in Motor Insurance Industry*. In: Prasad B. (eds) *Soft Computing Applications in Business. Studies in Fuzziness and Soft Computing*, vol 230. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-79005-1_7
18. Zaquieu, J.R. (2019). *Customer clustering in the health insurance industry by means of unsupervised machine learning*.

Додатки

Додаток 1. Розподіл значень змінних до та після корекції викидів.



Додаток 2. Описова статистика змінних у 2019 р.

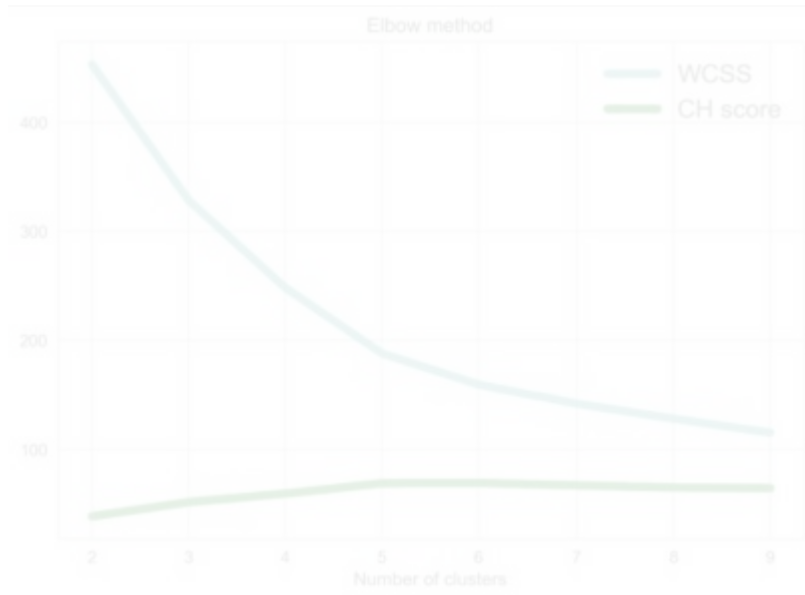
Додаток 2.1. Описова статистика змінних моделі.

	ROA	Offices	Re-to-premiums	% mandatory premiums	Corporate
Середнє	0.10	6.26	0.17	0.49	0.13
Ст. відхилення	0.25	18.44	0.25	0.34	0.27
Мін	-0.22	0.00	0.00	0.00	0.00
Q(25%)	0.02	0.00	0.00	0.18	0.00
Q(50%)	0.05	0.00	0.03	0.44	0.01
Q(75%)	0.09	0.00	0.29	0.83	0.08
Макс	2.19	115.00	0.90	1.00	1.00

Додаток 2.2. Описова інформативних показників компаній.

	Re-to-provisions	Loss ratio	Mean premium	Wages/Premiums	Concentration
Середнє	0.23	0.49	243.71	0.02	0.70
Ст. відхилення	0.25	0.50	1199.14	0.02	0.18
Мін	0.00	-0.63	0.00	0.00	0.33
Q(25%)	0.03	0.09	0.79	0.00	0.57
Q(50%)	0.14	0.43	2.69	0.01	0.69
Q(75%)	0.37	0.73	26.32	0.02	0.81
Макс	1.72	3.33	13768.22	0.07	1.00

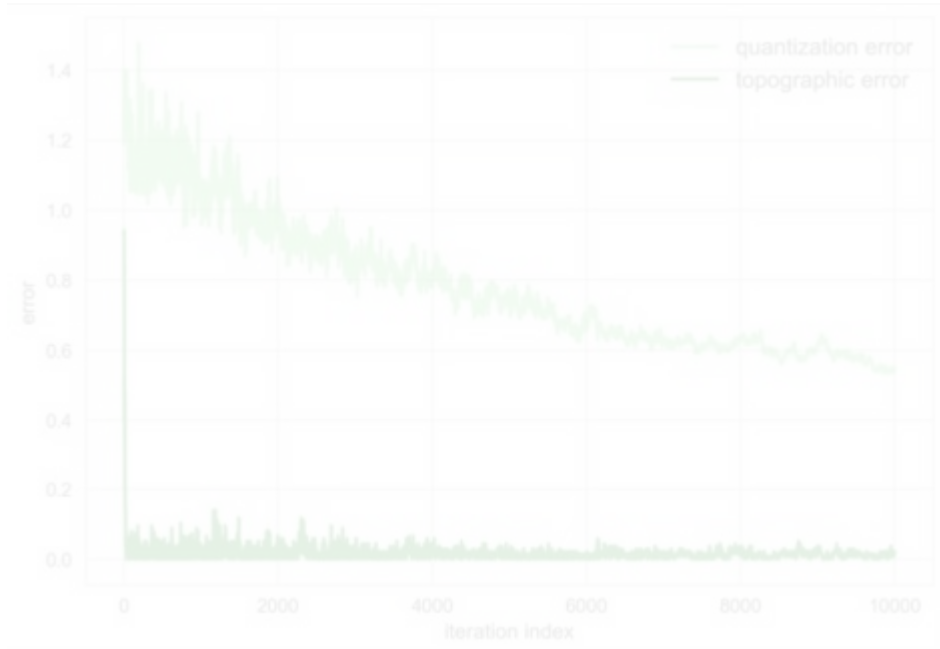
Додаток 3. Критерії вибору кількості кластерів.



Додаток 4. Кількість активацій нейронів на карті Кохонена.



Додаток 5. Динаміка навчання мережі Кохонена.

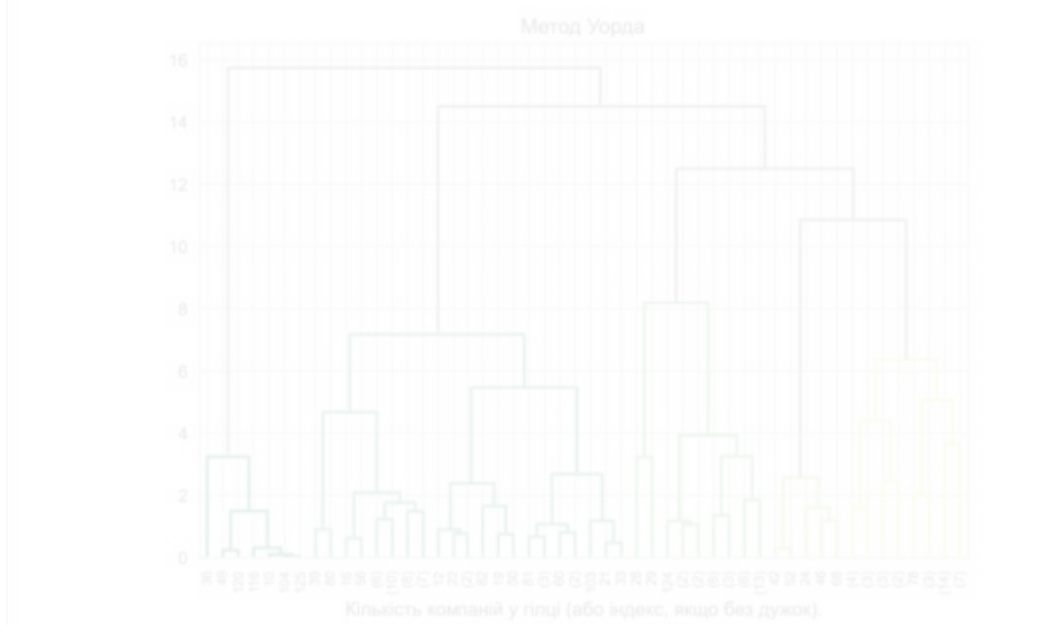


Додаток 6. Результати методу Уорда.

Додаток 6.1. Середні значення характеристик та кількість компаній кластерів.

	ROA	Offices	% mandatory premiums	Corporate	Re-to-premiums	Кількість компаній (2020)
0 - Роздрібна	0,046	0,143	0,044	0,207	0,009	35
1 - Універсальна "великі"	0,057	92,8	0,278	0,487	0,001	5
2 - Універсальна "малі"	0,002	6,838	0,537	0,393	0,066	37
3 - Корпоративна	0,034	5,269	0,062	0,847	0,026	52
4 - Перестрахова	-0,032	0	0,001	0,096	0,849	7

Додаток 6.2. Дендрограма.



Додаток 7. Результати методу К-медоїд.

Додаток 7.1. Координати центрів кластерів.

	ROA	Offices	% mandatory premiums	Corporate	Re-to-premiums
0 - Роздрібна	0,002	0	0	0,242	0,004
1 - Універсальна "великі"	0,067	44	0,242	0,433	0,017
2 - Універсальна "малі"	0,001	0	0,663	0,342	0,068
3 - Корпоративна	0,011	0	0,001	0,913	0,004
4 - Перестрахова	0,001	0	0	0,038	0,893

Додаток 7.2. Кількість компаній, що увійшли у кластери у досліджуваний період.

	2019	2020
0 - Роздрібна	44	40
1 - Універсальна "великі"	14	14
2 - Універсальна "малі"	28	28
3 - Корпоративна	52	46
4 - Перестрахова	14	8

Додаток 8. Опис кластерів

Додаток 8.1. Координати центрів кластерів.

	ROA	Offices	% mandatory premiums	Corporate	Re-to-premiums
0 - Роздрібна	0.030	0.35	0.050	0.242	0.015
1 - Універсальна "великі"	0.065	62.33	0.277	0.482	0.008
2 - Універсальна "малі"	0.014	6.14	0.637	0.368	0.057
3 - Корпоративна	0.027	1.19	0.047	0.887	0.027
4 - Перестрахова	-0.020	0	0.001	0.141	0.810

Додаток 8.2. Додаткові описові характеристики кластерів (2020 р.).

	Re-to-provisions	Loss ratio	Mean premium	Wages/Premiums	Concentration
0 - Роздрібна	0.186	0.457	5.424	0.065	0.759
1 - Універсальна "великі"	0.250	0.381	1.669	0.060	0.466
2 - Універсальна "малі"	0.168	0.390	41.90	0.064	0.637
3 - Корпоративна	0.275	0.225	254.20	0.051	0.701
4 - Перестрахова	0.119	0.049	303.97	0.001	0.679

Додаток 8.3. Кількість компаній, що увійшли у кластери у досліджуваний період.

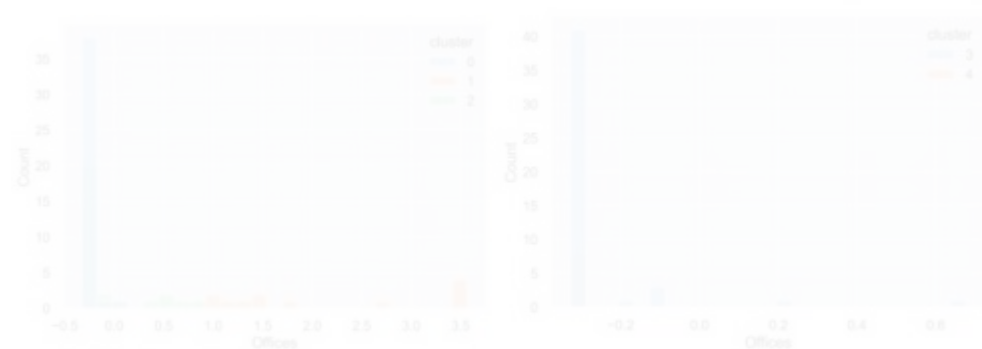
	2019	2020
0 - Роздрібна	42	40
1 - Універсальна "великі"	13	12
2 - Універсальна "малі"	29	29
3 - Корпоративна	52	47
4 - Перестрахова	15	8

Додаток 9. Гістограми ознак виділених кластерів (стандартизовано).

Додаток 9.1. ROA.



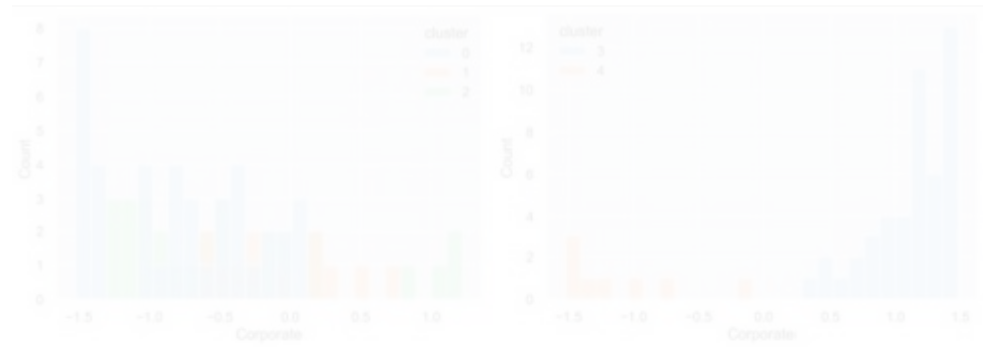
Додаток 9.2. Offices.



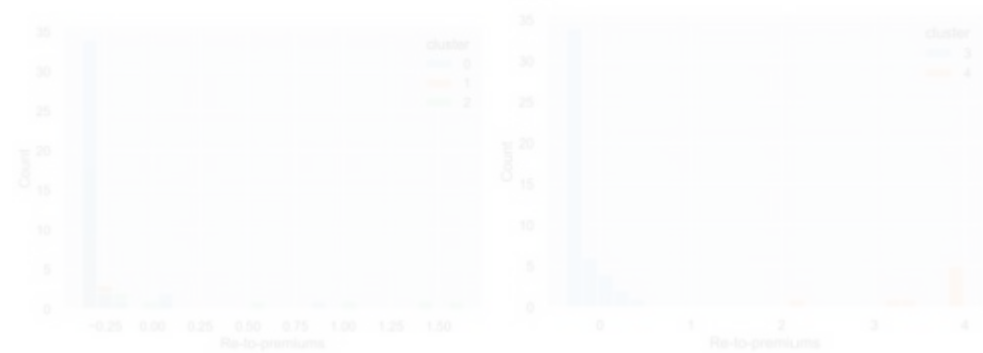
Додаток 9.3. % mandatory premiums



Додаток 9.4. Corporate



Додаток 9.5. Re-to-premiums



Додаток 7. Евклідова відстань між нейронами (нормалізовано).



Схожість

Джерела з Інтернету

143

1	https://link.springer.com/chapter/10.1007/978-3-540-79005-1_7	30 джерел	0.66%
2	https://samafind.sama.gov.sa/EDS/Search?dfApplied=1&filter%5B%5D=EXPAND%3A%22fulltext%22&filter%5B%5D=L	14 джерел	0.4%
3	https://repub.eur.nl/pub/17731	6 джерел	0.36%
4	https://www.semanticscholar.org/paper/Clustering-Austrian-Banks%E2%80%99-Business-Models-and-Peer-Ferstl-Ser	2 джерела	0.33%
5	http://www.investigo.biblioteca.uvigo.es/xmlui/bitstream/handle/11093/1263/ValladaresRodr%c3%adguez_SoniaMar	20 джерел	0.3%
6	http://www.immsp.kiev.ua/postgraduate/Biblioteka_trudy/ShtuchnNejronMeregNester2004.pdf	7 джерел	0.26%
7	https://coek.info/pdf-belief-in-karma-how-cultural-evolution-cognition-and-motivations-shape-belief-in.html	24 джерела	0.26%
8	http://hdl.handle.net/10366/148435	24 джерела	0.23%
9	https://core.ac.uk/download/pdf/215523128.pdf	2 джерела	0.22%
10	http://thesesups.ups-tlse.fr/5361/1/2022TOU30106.pdf	6 джерел	0.22%
11	https://researchr.org/publication/KramaricBDZZ18		0.19%
12	https://www.semanticscholar.org/paper/Customer-clustering-in-the-health-insurance-by-of-Zaqueu/8407767663b3b607415e89...		0.18%
13	https://dl.acm.org/doi/10.5555/1283383.1283474		0.18%
14	https://www.semanticscholar.org/paper/Fraud-Prevention-in-the-Leasing-Industry-Using-the-Bach-Vlahovic/633a385fe0dabc2a8		0.18%
15	https://ideas.repec.org/a/onb/oenbfs/y2021i42b1.html		0.15%
16	https://uk.wikipedia.org/wiki/%D0%94%D0%B8%D1%84%D0%B5%D1%80%D0%B5%D0%BD%D1%86%D1%96%D0%BC	2 джерела	0.12%
17	https://en.wikipedia.org/wiki/Maximum_flow_problem		0.11%